Eur J Appl Physiol (2011) 111:1917–1927
DOI 10.1007/s00421-011-1828-0

ORIGINAL ARTICLE

Recognition of activities in children by two uniaxial accelerometers in free-living conditions

N. Ruch · M. Rumo · U. Mäder

Received: 28 June 2010 / Accepted: 3 January 2011 / Published online: 20 January 2011

Abstract The aim of this study was to develop a classification procedure for accelerometer data to recognize the mode of children's physical activity (PA) in free-living conditions and to compare it with an established cutoff method. Hip and wrist accelerometer data with an epoch interval of 1 s were collected for 7 days from 24 girls (age: 10.7 ± 1.7 years) and 17 boys (age: 10.6 ± 1.6 years). Videos were recorded during the same 7 days at several points of time at school and during leisure time. Each second of video data was labeled as one of nine activity classes. A classification procedure based on pattern recognition algorithms was trained with the accelerometer data relating to respective video labels of half of the children and tested against the data from the other half of the children. The overall recognition rate of the classification procedure was 67%. The procedure was able to classify 90% of stationary activities, 83% of walking, 81% of running and 61% of jumping activities. The remaining activities could not be recognized by the main classifier. This study developed a classification procedure based on well-accepted accelerometers and video recordings to recognize children's PA in free-living conditions. It has been shown to be valid for the activities of being stationary, walking, running and jumping. In contrast to former measurement and analysis procedures, this method is able to determine the modes of specific activities among children. Consequently, the presented classification procedure provides additional information on the PA behavior in children registered by established accelerometers.

Keywords Classification procedure · MTI · Youth · Motion sensor

Introduction

Regular physical activity (PA) is considered to be an important aspect for a healthy lifestyle in children (Riddoch et al. 2004; Fuchs et al. 2001). However, there is still a lack of quantifiable conclusions regarding PA and its possible health effects in children. This might be caused by the general difficulty in measuring PA in this population. Most existing questionnaires are not recommended for distribution to children due to their lack of cognitive ability to accurately recall their PA behavior (Baranowski 1988; Kohl et al. 2000). The technique of direct observation needs substantial time efforts to measure PA (McKenzie 1991; McKenzie et al. 1991) and interference of the observers with children was reported (Bailey et al. 1995). For some time, accelerometers have been widely used to assess PA in children. These devices provide a simple, low-cost method to measure intensity, duration and frequency of activities in children and are well accepted in this population (Eissa et al. 1999; Janz 1994). Linear regressions between vertical (Freedson and Pober 2005; Trost et al. 2000; Puyau et al. 2002; Ekelund et al. 2004) and tri-axial (Tanaka et al. 2007) acceleration output during different activities and physiological variables (e.g., VO_2 or METs) were used to determine ranges of accelerometer output corresponding to different intensity levels of PA. However, vertical accelerations were not found to develop linearly at high velocities (Brage et al. 2003). Several

Communicated by Klaas R Westerterp.

N. Ruch (✉) · M. Rumo · U. Mäder
Physical Activity and Health Unit, Swiss Federal Institute of Sport SFIS, Hauptstrasse 247, 2532 Magglingen, Switzerland
e-mail: nicole.ruch@baspo.admin.ch

studies have reported mean vertical hip counts with widely used devices for different activities in children (Puyau et al. 2002; Treuth et al. 2004), but these studies did not focus on the recognition of the mode of activities. Tanaka et al. (2007) showed that stationary activities, walking and running could visually be distinguished by the hip acceleration counts; however, their main aim was not activity recognition. Recently, methods using pattern recognition approaches were developed to classify accelerometer and other objectively collected data into several activity classes (Bao and Intille 2004; Pober et al. 2006; Bonomi et al. 2009). These authors calculated different features over pre-defined time windows from various sensors and used various classifiers to discriminate between activities. However, one of these studies used a very burdensome setup consisting of 17 different sensors that were not appropriate for children (Pärkkä et al. 2006). Activities were often measured in laboratory conditions (Bonomi et al. 2009; Pober et al. 2006) or in unsupervised conditions where participants followed a scenario (Bao and Intille 2004; Pärkkä et al. 2006). Therefore, data were not collected in free-living conditions. As activities were continuous and some of them adult specific, they did not reflect children's activities. The setups and protocols of these studies could not be transferred directly for use with children despite the promising recognition results of over 80% in most of the classifiers used in these studies. Consequently, a pattern recognition procedure to recognize various children-specific activities with simple devices that are highly accepted in this population is yet to be developed. Therefore, the aim of the present study was to compile and validate a classification procedure that allows long-term data collection in free-living conditions and determines the mode of children-specific activities. Furthermore, accelerometer data of the assigned activity will be compared to established cutoff methods.

Methods

Subjects

A total of 24 girls and 17 boys were recruited from three suburban elementary schools. Responsible teachers were asked to distribute an information letter to the families of their pupils that invited them to participate in the study. A letter of informed consent was signed by parent and child before the child was included in the study. The study was approved by the local ethics committee. The participating children were randomly assigned to the training data set group (TRG) ($n = 21$) and testing data set group (TEG) ($n = 20$) to provide the basis for one validated final model that allows further analysis of the PA behavior of children.

Table 1 Characteristics of participating children randomly split up into a training data set group (TRG) and a testing data set group 9

	TRG ($n = 21$) (12♀, 9♂)	TEG ($n = 20$) (12♀, 8♂)
Age (years)	♀ 10.8 ± 1.3	♀ 10.6 ± 0.8
	♂ 10.5 ± 1.3	♂ 10.9 ± 1.0
Height (m)	♀ 1.5 ± 0.1	♀ 1.5 ± 0.1
	♂ 1.5 ± 0.1	♂ 1.5 ± 0.1
Weight (kg)	♀ 38.5 ± 10.2	♀ 37.6 ± 6.9
	♂ 41.1 ± 10.3	♂ 38.0 ± 6.1
PA (counts/min)	♀ 573 ± 151.2	♀ 526.2 ± 77.5
	♂ 570.5 ± 121.3	♂ 675.5 ± 88.0

Data from the TRG were used to train the classification system, and data from the TEG were used to evaluate it. Subject data are shown in Table 1.

Measurement procedures

As accelerometer data were collected in a natural environment in this study, it was assumed that arm activities would occur quite often. Therefore, children were asked to wear two accelerometers, one at their wrist and one at their hip for 1 week. During the measurement week, three to four bouts of 1–3 h (total recording time: 7.3 ± 1.7 h/child) of the children's activities were recorded on a video system. Recordings were taken at ordinary school during classes (49.4 ± 9.4% of recording time), physical education (8.9 ± 4.1%), during unstructured playing at home indoors (21.4 ± 14.1%), outdoors (11.1 ± 11.7%) and during structured leisure time activities (9.9 ± 7.5%). Each day of the week, an activity log was filled in by children with the help of their parents.

Accelerometers

Although a previous study (Tanaka et al. 2007) showed that synthesized tri-axial accelerometer data of stationary, walking and running activities can be visually distinguished in a graph, uniaxial accelerometer counts were used for the recognition procedure, as they are currently the devices that are used in European PA monitoring studies (Andersen et al. 2006). Accelerometers used in the present study (GT1M, The Actigraph, FL, USA) had been validated earlier (Janz 1994; Melanson and Freedson 1995; Trost et al. 1998). They demonstrated good intra-instrument reliability (Metcalf et al. 2002) and showed the lowest amount of variance when compared with other activity monitors (Welk et al. 2004). They have been found to be well accepted in children (Janz 1994; Eissa et al. 1999). Furthermore, several cutoff points for these devices were established (Puyau et al. 2002; Ekelund et al. 2004; Freedson and Pober 2005) allowing the comparison between them and the present classification system. The accelerometers record 30 measurements per second and integrate these values continuously over time. Activity counts are the sum of the accelerations measured over a selected period (epoch time). For the present study, it was set to 1 s with respect to spontaneous, intermittent activity behavior in children (Bailey et al. 1995). These authors reported that the median of moderate (vigorous) activity in children was 6 s (3 s). As their sampling rate was 3 s, it was supposed that activities in children might be even shorter. Therefore, setting a time window of several seconds for feature calculation would not be children specific. Furthermore, a classification method that uses acceleration data with unit 'counts' is desired to keep the complexity of the model low and to allow comparison to former cutoff methods.

Video recording system

During the measurement week, places of recordings were visited to position the video cameras unknown to the children in order that their activity behavior was not influenced. The behavior of children was recorded with three digital observational video cameras (Wireless camera 830G, Lupus Electronics, Landau, Germany). These cameras were very small (1.5 cm × 3 cm × 3 cm). Video recordings were stored on a hard disk recorder (DVR Client Manager, RV100 Series, Lupus Electronics, Landau, Germany). If parents reported that their child changed their indoor or outdoor playing sites often, a portable camera on a stand was left at the child's home to be placed by the parents at the playing sites. If this was not possible, a researcher followed the child from a distance with a portable camera.

Data processing and analysis

As it was the aim of this study to recognize activities in a most natural environment of children, video observation was considered the most appropriate method to determine the mode of activity. Video sequences were analyzed with a software (Dartfish Team Pro 4, Dartfish, Fribourg, Switzerland), which indicated the time of the recording and offered a function to label video sequences as activity categories. Activity categories were based on the category system of Bailey et al. (1995) and pre-tested in a sample of the recordings of the TEG (4 h in total of a sample of recordings from randomly chosen 10 children), as was done previously in another observational study (Bailey et al. 1995). Further activity categories were added if necessary. Activity classes chosen for the present study were stationary activities, walking, running, jumping, scooter, floor exercise, biking, horseback riding and crawling (Table 2). The random test set of video sequences was further used to train all of the observing researchers on how to properly label the video data. The Kappa coefficient for inter-observer reliability (0.90–0.91) was comparable to other observational studies (Epstein et al. 1984; Bao and Intille 2004). The video recordings were labeled with an accuracy of a millisecond and rounded off to the next higher second. Transitions of activities were labeled with the more strenuous activity class before or afterward. With the help of the time lines of the video recordings and the accelerometer data, both data sets were then synchronized to relate the video labels with the accelerometer values. During video analysis, researchers were asked to mark

Table 2 The nine activity categories chosen according to the most frequent activities in the video recordings and the description of activities that were assigned to the respective category

Activity category	Description of activities classified in this category
Stationary activities	Activities requiring the person to remain in the same place, such as lying, sitting, standing, kneeling
Walking	All velocities of walking, walking up- or downhill, climbing or descending stairs, walking while playing with a ball
Running	All velocities of running, running up- or downhill, running on stairs, running while playing with a ball
Jumping	Single jumps while playing, rope skipping, jumping while playing ball games (basketball, tennis, etc.), jumping down from an object, jumping onto an object
Floor exercise	Somersault, handstand, falling over or down
Biking	All velocities of riding a bike with pedaling (biking without pedaling was classified as stationary activity)
Horseback riding	Striding, trotting, galloping (acrobatics on the horse were classified as floor exercise)
Crawling	Crawling, dynamic stretching, rolling on the floor
Scooter	All velocities of riding a scooter (standing, kneeling or sitting on the scooter without kicking was classified as stationary activity)

homogenous sequences of activities. These were sequences where children clearly performed one of the activities of the category system. Only these were used to train the classifiers.

Accelerometer data of the TRG and their respective activity labels were used to train the classification procedure. In our data set, stationary activities produced multiple data points consisting of the same values. As this minimized the variance used by the parametric classifiers, we decided to delete all but one data points of the same value. The goodness of the classification procedure to recognize most natural activity data was tested with data from the TEG that contained not only clear sequences, but also transitions and sequences where combined activities were performed (e.g., walking while tossing a ball). Multiple data points of the same value were left in the data of the TEG. Activity classes found during video analysis in the TEG were only integrated into analysis if more than 50 accumulated seconds of this activity existed. Furthermore, the number of data points in all activity classes was reduced to the size of the smallest class for each child separately to give each class the same weight within a child. Comparison of video labels and labels found by the classification system in the data of the TEG resulted in proportional recognition rates to picture the quality of the classification procedure. Mean, minimum and maximum recognition rates of all children of the TEG are demonstrated in the results.

Hip accelerometer data of the TEG were also classified by the cutoff method. A variety of cutoff points are available (Ekelund et al. 2004; Freedson and Pober 2005; Puyau et al. 2002), but up to now it has remained unclear which of these provided the most valid outcome (Bassett 2007). Cutoff points in the present study were chosen according to the most varied and natural activities used during their development (Puyau et al. 2002). Cutoff points for the hip data were <800, <3,200, <8,200 and ≥8,200 counts for sedentary, light, moderate and vigorous activity, respectively. They were divided by 60 (1-s cutoff points) to analyze hip accelerometer on a second-by-second basis to allow a comparison to the recognition results of the present study. Although cutoff points may not decrease linearly when analyzing data with a lower epoch time, this was considered to be the best approach to compare the outcome of the cutoff method with the classification procedure in the present study.

The classification process combined three different classifiers such as k-nearest neighbor (k-NN) (Edwart and Fischer 1970), normal density discriminant function (NDDf) (Duda et al. 2001) and a custom decision tree (CDT) to overcome inherent limitations of the single classifiers. The second classifier, k-NN, assigns a test data point to the class that most of its nearest training data points belong to. This classifier is dependent on the original data set, which can be provided for future studies. In the present study, a data point was classified by a majority vote of its 111 'nearest neighbor' data points to the respective activity class. Its height was determined by calculating the error rates for different numbers of neighboring points. The NDDf uses the class conditional parameters of the normal distribution for each class j in the training data (Table 1) to determine the discriminant function for each data point x to be tested:

$$g_j(x) = -\frac{1}{2}(x-\mu_i)\cdot\Sigma^{-1}\cdot(x-\mu_i) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(\det\Sigma) + \ln(p_j), \qquad (1)$$

where g is the discriminant function of activity j, d is the number of data points and p_j is the a priori probability of j.

The CDT was based on cutoff points that were chosen visually by the researcher (Fig. 1). As a single classifier may valorize one activity class over the others, the use of meta-classifiers may decrease the weight of the decision of a single classifier. Consistent with previous research, the most successful meta-classifier (Ravi et al. 2005), a majority vote (MV), was used that determined the major class. Therefore, the final decision of the MV was based on the decision of all three classifiers. Data that were categorized differently by each classifier were annotated as 'not assigned'. As it was not clear if wrist data were necessary for good recognition, the whole classification procedure was repeated using only hip acceleration counts. Borders of the CDT were specifically adapted (Fig. 2). All classification procedures were done using Matlab 5.3 (Mathworks, NM, USA). The remainder of the statistical analysis was done using SPSS 15 (SPSS Inc., Chicago, IL, USA).

Results

Recognition rates of the classification procedure

Test data contained 2,587 s of stationary, walking and running activities, 2,228 s of jumping activities, 168 s of floor exercise, 395 s of biking, 291 s of horseback riding and 783 s of crawling. Descriptions of the training data are given in Table 3.

As the activity class 'scooter' was only performed in the TRG, results include recognition rates for the remaining eight activities (stationary, walking, running, jumping, floor exercise, biking, horseback riding, crawling) only.

Recognition rates were best for the k-NN and the MV and lower in the CDT and NDDf (Table 4). The recognition rates of all classifiers were higher when wrist accelerometer counts were included in the classification procedure. k-NN was the only classifier that was able to

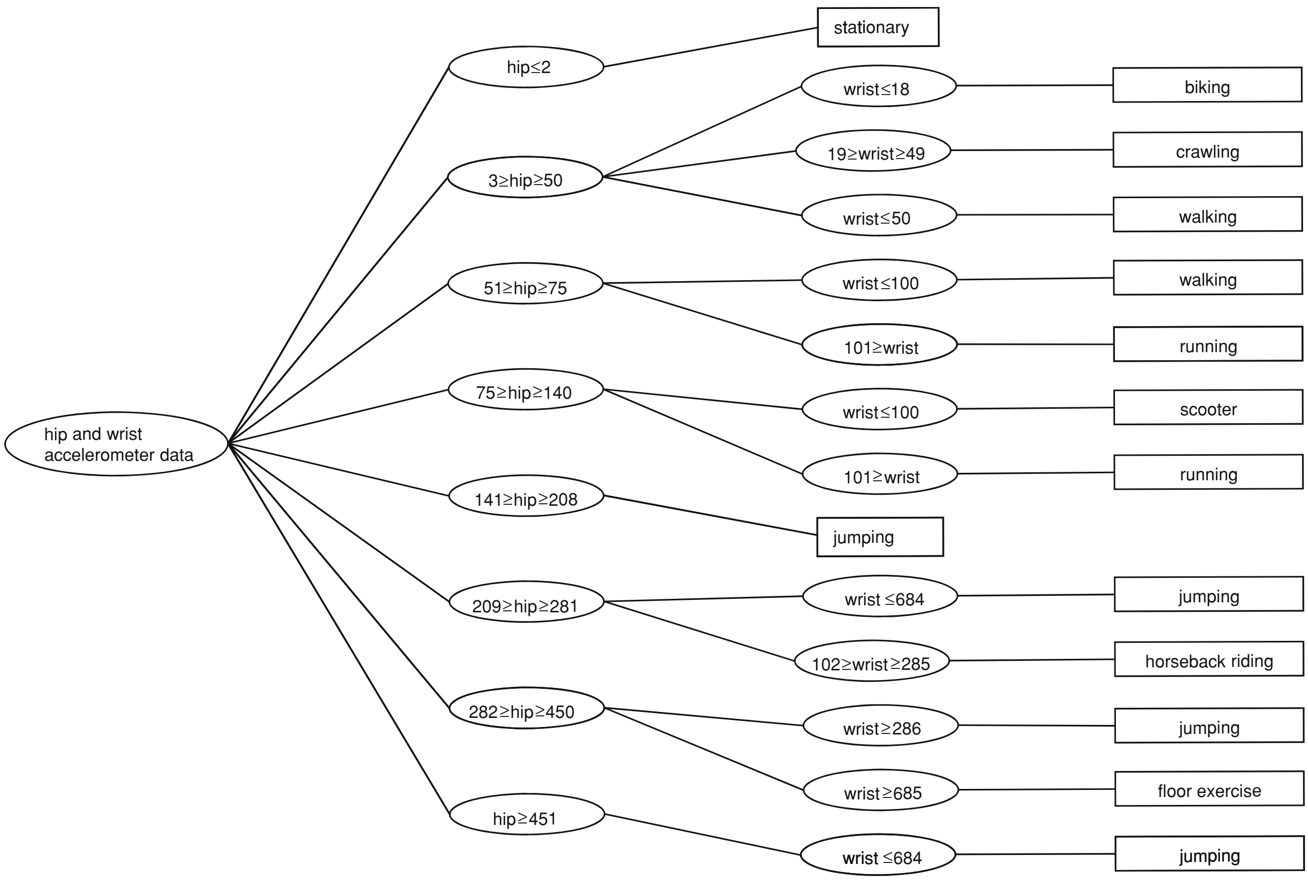

Fig. 1 Custom decision tree using hip and wrist accelerometer data. *Numbers* represent the threshold values in counts s^{-1}

classify more than 50% when only hip data were used. The estimated classifications of predicted activities (columns) of the k-NN, NDDf, CDT and MV for each of the activity classes using hip and wrist acceleration values are shown in Fig. 3. Floor exercises, biking, horseback riding and crawling were only assigned by the CDT. Stationary activities could not be recognized by the NDDf (0%), but were registered by the k-NN to the most part (95%). NDDf resulted in the highest recognition rate for walking (94%). Running was best recognized by the CDT (84%) and jumping by the k-NN (79%). All other activities were best recognized by the CDT (floor exercise: 2%, biking: 23%, horseback riding: 1%, crawling: 13%).

Comparison of results from the cutoff method to classification results

The proportions of cutoff-based intensity levels during the activities of the classification procedure are shown in Table 5. Most of the stationary data found by the classification procedure were assigned to the sedentary class by the cutoff method. Walking was assigned mostly to light activities, running mostly to moderate activities and all jumping data were classified as vigorous activities. Data categorized as 'not assigned' by the classification procedure were assigned to sedentary and low activities mostly, to moderate activities to a smaller part, and to vigorous the least.

Discussion

Overall recognition rates of the different classifiers

The recognition rate of the MV and the k-NN in the present study were 67%. Other authors reached 83–84%, 82–86% and 90.4–93.1% (Bao and Intille 2004; Pärkkä et al. 2006; Bonomi et al. 2009). It is supposed that the high recognition rates were generated by collecting the data under supervised laboratory conditions (Ravi et al. 2005), in an obstacle course (Bao and Intille 2004) or during an unsupervised scenario (Pärkkä et al. 2006) that was followed by the subject. Therefore, these studies did not measure activities during daily life, but in structured laboratory

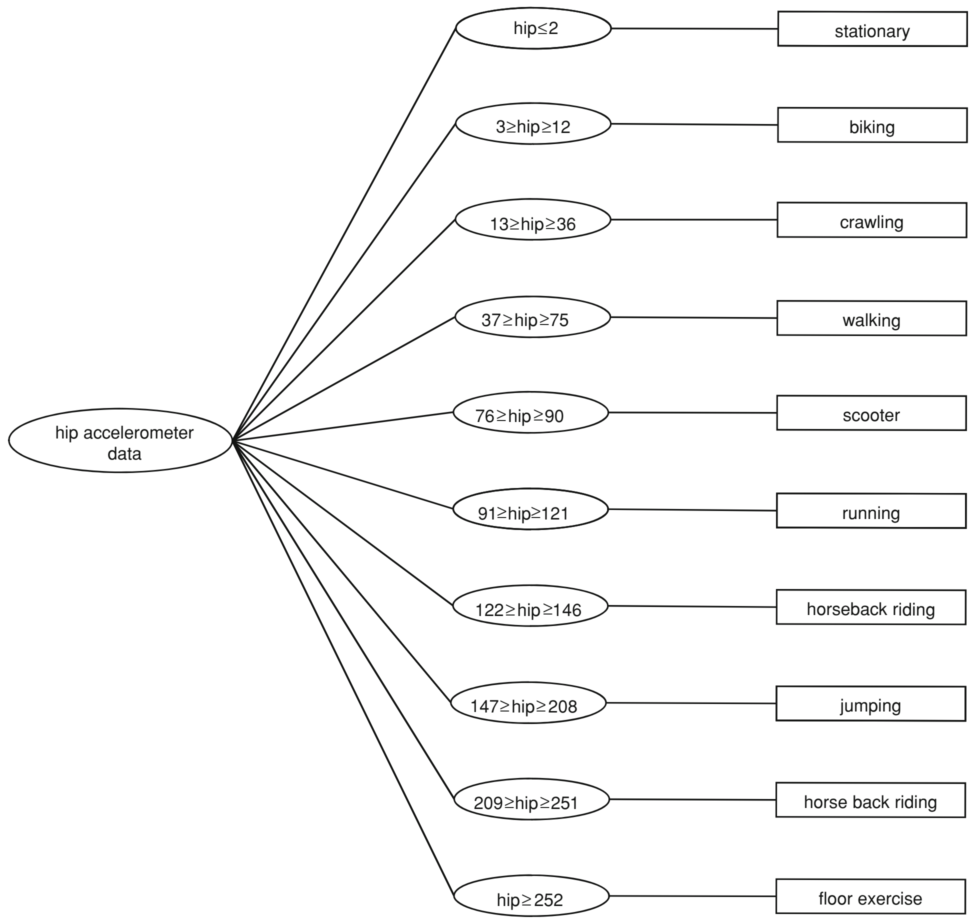

Fig. 2 Custom decision tree using hip accelerometer data only. *Numbers* represent the threshold values in counts s^{-1}

Table 3 Mean, standard deviation and covariance of hip and wrist accelerometers in the TRG

Activity	Hip	Wrist	Cov hip–wrist
Stationary	18.9 ± 6.7	115.6 ± 10.8	−35.80
Walking	36.6 ± 7.9	141.7 ± 36.3	7.50
Running	95.3 ± 10.5	337.5 ± 59.6	104.70
Jumping	213.7 ± 24.7	335.2 ± 55.8	74.30
Floor exercise	262.7 ± 35.1	314.9 ± 60.8	875.40
Biking	36.9 ± 26.7	62.0 ± 44.9	677.90
Horseback riding	149.0 ± 30.3	155.0 ± 33.6	558.70
Crawling	39.7 ± 30.6	69.5 ± 51.1	142.00

Cov hip–wrist covariance of hip and wrist data

conditions where activities were performed by adults providing long and clear sequences of an activity. The aim of the present study was to measure activities in daily life as in epidemiological studies to collect representative data. Therefore, data to train and test our classifiers were collected in a non-structured and most natural daily life setting, producing higher intra-class variance in the accelerometer data than when data were collected in laboratory conditions. This may explain the lower recognition rates compared to the previous studies.

Table 4 Mean, minimal and maximal recognition rates of the different classifiers using only hip and combined hip and wrist data in the TEG

	Hip data (%)	Hip and wrist data (%)
k-NN ($k = 111$)	59 (40/74)	67 (51/85)
NDDf	22 (14/33)	49 (35/62)
CDT	48 (34/90)	64 (48/85)
MV	44 (28/68)	67 (51/85)

Other authors used a setup collecting multi-dimensional acceleration data on either one (Ravi et al. 2005; Bonomi et al. 2009), two (Bao and Intille 2004; Pärkkä et al. 2006) or various (Bao and Intille 2004) locations of the body. This might also have increased the recognition rates

Fig. 3 The proportion of estimated classifications (legend) per predicted activity mode (100% of the specified activity found during video analysis in each *bar*) for stationary, walking, running, jumping, floor exercise, biking, horseback riding and crawling activities. Results of the k-nearest neighbor (k-NN), normal density discriminant function (NDDf), and the custom decision tree (CDT) classifiers are shown in *different columns*. The fourth column shows the results of the majority vote between the classification results of the three classifiers (MV) ($n = 20$) ▶

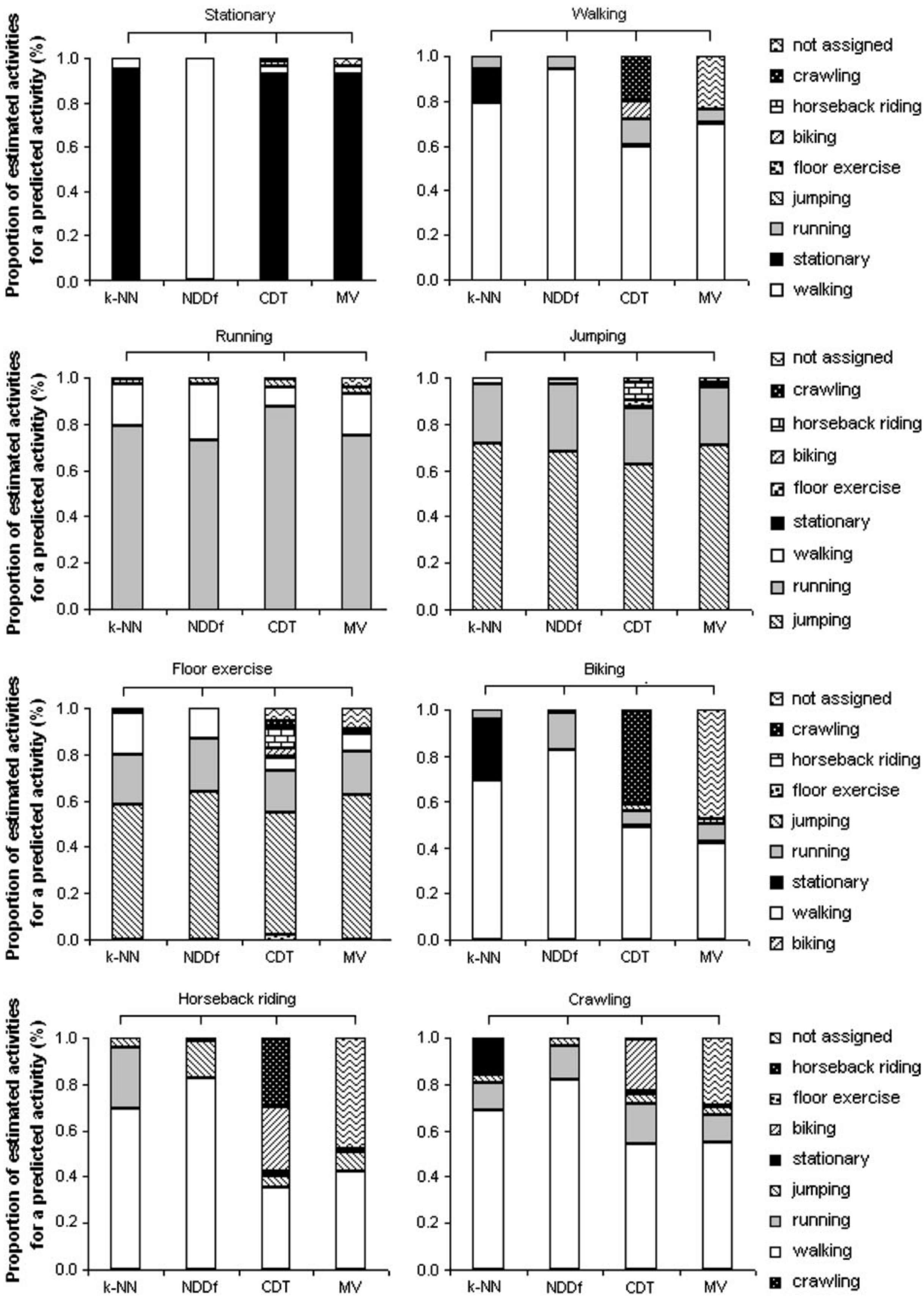
Stationary
Walking
Running
Jumping
Floor exercise
Biking
Horseback riding
Crawling
Proportion of estimated activities for a predicted activitiy (%)
1.0
0.8
0.6
0.4
0.2
0.0
k-NN
NDDf
CDT
MV
not assigned
crawling
horseback riding
biking
floor exercise
jumping
running
stationary
walking

Table 5 Confusion matrix of the classification results versus the results of the cutoff method

Activities assigned by the cutoff method	Activities assigned during the classification procedure				
	Stationary (%)	Walking (%)	Running (%)	Jumping (%)	Not assigned (%)
Sedentary	99.9	17.0	0	0	37.4
Low	0	58.9	1.5	0	39.6
Moderate	0	24.1	83.1	0	16.3
Vigorous	0.1	0	15.4	100.0	6.6

The *i*th row and *j*th column of this table contains the percentage of activity *j* found during the classification procedure that was assigned to intensity level *i* by the cutoff method

compared to our study. However, in one of the above studies, single classifiers produced lower recognition rates, such as 47–52% (Bao and Intille 2004), suggesting the combination of different classifiers in one study. The present study wanted to use established devices that are most simplified, as children cannot be expected to wear a complicated device for a long period. Nevertheless, the device should provide the possibility for long-term measurements (at least 7 days) as such periods are measured by PA monitoring studies (Andersen et al. 2006). The devices used in this study were suitable, even though they provided less information than other more sophisticated devices. Therefore, the present recognition rates can be seen as adequate for a study classifying, on the one hand, the most natural data of free-living children and, on the other hand, using a most simple device that is known to be well accepted by children.

The only study to our knowledge that performed an MV was by Ravi et al. (2005). They found that an MV (90.6–99.6%) performed best of all the tested single and meta-classifiers. In contrast, the present study found no difference between the best single classifier (k-NN) and the MV, raising the question why to perform the two worst single classifiers. The advantage of the MV is the recognition of activities as 'not assigned'. For example, crawling was to a great part classified as 'not assigned' by the MV. In contrast, the k-NN annotated crawling to stationary or walking. Therefore, the MV can provide additional information in terms of identifying unknown (other than stationary, walking, running and jumping) activities.

In the present study, TRG and TEG contained different subjects to make sure the algorithm can be used for any other subject. Ravi et al. (2005) found that the recognition rate decreased to 65% if the classifier was tested with data from different subjects than it was created. Therefore, recognition rates of studies using multi-dimensional accelerometer data might provide similar data to the present study, if they try to generalize their classification procedure for any other subjects.

Recognition rates of the single activity classes

Recognition rates of the MV of some specific activities that were observed very often during video analysis were moderate (walking/running/jumping: 69.1/75.2/70.9%) to high (stationary: 92.6%) (Fig. 2). With their best classifiers, other studies (Bao and Intille 2004; Pärkkä et al. 2006) provided results similar to the present study with regard to lying down (87–95%) and sitting/standing (94.8–96%). Bonomi et al. (2009) found very high recognition rates for lying down (100%) and similar recognition rates for sitting (87.4%), but lower values for standing (62.4%). Consequently, it is suggested that stationary activities can generally be discriminated from active behavior. Running was recognized more effectively (89.7–100%) by other authors (Bao and Intille 2004; Pärkkä et al. 2006; Bonomi et al. 2009) than in the present study, where wrongly assigned running data were mostly classified as walking. This is likely to be attributed to a smooth transition from walking to running in children and therefore to difficulties in labeling and discriminating between these two activities. This is supported, as the same authors found better results for the recognition of walking (78–99.5%) than in the present study. Furthermore, activity classes in the present study included all transitions that occurred. This might have caused greater, but more natural variation within the data compared to the predefined activities of the other studies (Bao and Intille 2004; Pärkkä et al. 2006; Bonomi et al. 2009). These variations are likely to have contributed to the slightly lower recognition rate of the classifiers in the present study.

The MV had low recognition rates in activity classes that were performed only by a small number of children in the TEG (scooter, crawling, horseback riding, floor exercise) (Fig. 2). The scooter activity occurred rarely and no child in the test group provided enough data for analysis. Crawling and cycling were mostly classified as walking. As floor exercise was defined as a class containing high-impact activities on the floor, most data of this class were annotated to jumping activities. Accelerometer data of horseback riding was mostly classified as walking (27.0%) or not assigned (48.4%). These limitations have to be considered in studies focusing on shifts from one activity to another. If the main focus is on energy expenditure, misclassification is not as serious, as misclassified activities produce similar energy expenditure to the actual activity (Arvidsson et al. 2007; Treuth et al. 2004), except biking that is underestimated in terms of energy expenditure when classified as walking. As biking, horseback riding, floor exercise and crawling activities were recognized as typical in children in the TRG during video analysis, these activities should be accounted for in future pattern recognition studies, although they were less frequently performed than walking, running, jumping and stationary activities.

Possible adaptations of the method and their influence on recognition rates

Recently, higher overall recognition rates than those of the present study were reported (quadratic discriminant function: 70.9%, hidden Markov model: 80.8%) using the same type of accelerometer in adults (Pober et al. 2006). The authors discriminated four activities that could not be discriminated on the basis of hip accelerometer data with unit 'counts' and an epoch period of 1 s but of different energy expenditures, such as uphill walking and walking over a level surface or desk work and vacuuming. The authors of that study extracted different features (mean and standard deviation) over respective data windows (15 s) from raw data (30 Hz) of the same devices as used in the present study. The same procedure has been performed in other studies using different devices (Bao and Intille 2004; Pärkkä et al. 2006); Bonomi et al. 2009). Pober et al. (2006) found comparable recognition rates to our study for walking (58.2–62.6%) and desk work (stationary activity) (97.3–100%). Similarly, Tanaka et al. (2007) showed that data of walking and stair climbing could be distinguished by a discriminant analysis using the ratio of horizontal and vertical acceleration of one device worn on the hip. Therefore, multi-dimensional accelerometers in combination with the use of feature selection could improve the recognition of activities that cannot be separated when only hip acceleration counts are measured. Pober et al. (2006) calculated features over 15 s. The device of Tanaka et al. (2007) obtained tri-axial acceleration every 40 ms that was averaged over 5 s. The duration of children's activities was reported to be 6 s (3 s) for moderate (vigorous) activities (Bailey et al. 1995). As it is suggested that these durations are even shorter as the sampling rate of this study was 3 s, the chosen length of the time windows of the previous studies (Pober et al. 2006; Tanaka et al. 2007) is not applicable to the analysis of data from children. In the present study, a high sampling rate (>1/s) would have been necessary if features were calculated for the chosen time windows (1 s). As data were collected over the long term (7 days) in the present study, storage capacity and battery life of the used accelerometers were not sufficient for such measures. Consequently, feature calculation over short, children-specific time windows and for multi-dimension devices would most likely improve recognition results in long-term studies. This will be possible for future studies as technology is developing. The present classification procedure was able to classify activity types based on acceleration counts measured at a 1-s epoch time, which was desired to reduce model complexity and to compare with former PA studies using cutoff points based on activity counts.

Comparison of results from the cutoff method to classification results

As certain activities were not recognized by the classification procedure (floor exercise, biking, horseback riding, crawling), cutoff results could be compared only to the recognition of the remaining activity classes. Most stationary activities found during the classification procedure were recognized as sedentary by the cutoff method. Walking was mainly classified as light activity and to a smaller degree as sedentary or moderate activity. Running was mainly recognized as moderate activity and to a smaller degree as vigorous activity. A great part of jumping was assigned to vigorous activity. These classifications reflect to a great part the similarity in the cutoff points of Puyau et al. (2002) and the CDT when only looking at the hip data for the respective classes. For example, Puyau et al. (2002) set the cutoff point for light activities at <3,200 (counts/min), and the CDT set the hip cutoff point for walking at 180–4,500 counts/min). The cutoff method provided useful information on the discrimination of activities in terms of their intensity level. However, it seems that it estimated intensity levels rather low. It is likely that this may have been caused in part by the cutoff points that were divided by 60 to account for the second-by-second data, as cutoff points may not decrease linearly when analyzing data with a lower epoch time. Furthermore, the cutoff points were developed by letting children perform continuous activities over a certain time period. The present study was developed on the basis of free-living activities that inherit greater variance and are short in duration. Therefore, the two systems focus on different quality of activity possibly causing the differences in the results. Accelerometer data were classified by the cutoff method into different intensity levels, whereas during the classification procedure, the same data were annotated to a single activity class. These differences indicate that there were large differences in how intense children performed a specific activity. As it was not possible to measure energy expenditure without limiting the demand concerning the inclusion of most natural free-living activities, it remains a challenge for future studies to relate different intensity levels of single activities to their respective energy expenditure.

Strengths and limitations of the study

The strength of this study is that daily activities in a most natural environment were measured to provide data for the TRG as well as for the TEG of the classification procedure. Devices were used that are well accepted in the measured population and used in European monitoring studies. However, the trade-off for the long-term measurement with these devices is the limitation of the sensor information that can be provided for this time period, as storage capacity and battery life are restricted. Another limitation is that some activities found in the TRG were underrepresented in the TEG (e.g., activity ‘scooter’). However, as we measured in a most natural environment, this represents reality as not all children perform the same activities. Videos were recorded only at set time points during the week, as this method is very time-consuming for both, collecting and analyzing the data. However, it is to our knowledge the only method that provides a detailed (on a second-by-second basis) insight into activities of children that are performed naturally in their usual environment.

Conclusion

The present study developed a classification procedure based on simple, well-accepted accelerometers to recognize children’s PA. As the present classification procedure was trained and tested with data collected in a most natural setting, and was based on the use of activity counts, an overall recognition rate of 67% can be regarded as adequate. The classification procedure has been shown to be valid for stationary, walking, running and jumping activities, but not for floor exercise, biking, crawling and horseback riding. The use of a meta-classifier such as a MV did not improve the recognition results, but provided additional information in terms of identifying unknown activities. When accelerometer data of the present study were classified by the cutoff method, data of the same mode of activity were classified into different intensity levels indicating that there were differences in how intense children performed an activity. In future, optimal measurement methods are required to recognize more specific activities that are yet to be integrated into the same activity category as in the present study. Within each activity category, the estimation of energy expenditure with multi-factor regression on the basis of objective activity measurements and subject-specific information is desired. The recognition of the mode of activities is the additional benefit of the present classification procedure in comparison to the cutoff method. It is therefore suggested that classification models as presented above could be followed in future studies on recognition of children-specific activities.

Acknowledgments The authors are grateful to the children and their parents for their willingness to participate in the study. The study was supported by grants from the Swiss Federal Council of Sports.

Conflict of interest The authors declare that they have no conflict of interest.

References

Andersen LB, Harro M, Sardinha LB, Froberg K, Ekelund U, Brage S, Anderssen SA (2006) Physical activity and clustered cardiovascular risk in children: a cross-sectional study (The European Youth Heart Study). Lancet 368:299–304

Arvidsson D, Slinde F, Larsson S, Hultén L (2007) Energy cost of physical activities in children: validation of SenseWear Armband. Med Sci Sports Exerc 39:2076–2084

Bailey RC, Olson J, Pepper SL, Porszasz J, Barstow TJ, Cooper DM (1995) The level and tempo of children’s physical activities: an observational study. Med Sci Sports Exerc 27:1033–1041

Bao L, Intille SS (2004) Activity recognition from user-annotated acceleration. In: Ferscha A, Mattern F (eds) Pervasive 2004. LNCS, vol 3001. Springer, Berlin/Heidelberg, pp 1–17

Baranowski T (1988) Validity and reliability of self report measures of physical activity: an information-processing perspective. Res Q Exerc Sport 59:314–327

Bassett DR Jr (2007) Validity and reliability issues in objective monitoring of physical activity. Res Q Exerc Sport 71:S30–S36

Bonomi AG, Goris AHC, Yin B, Westerterp KR (2009) Detection of type, duration, and intensity of physical activity using an accelerometer. Med Sci Sports Exerc 41:1770–1777

Brage S, Wedderkopp N, Franks PW, Andersen LB, Froberg K (2003) Reexamination of validity and reliability of the CSA monitor in walking and running. Med Sci Sports Exerc 35:1447–1454

Duda RO, Hart PE, Stork DG (2001) Pattern classification. Wiley-Interscience Publication, New York

Edwart AP, Fischer FP (1970) A generalized k-nearest neighbour rule. Inform Control 16:128–152

Eissa MA, Yetman RJ, Poffenbarger T, Portman RJ (1999) Comparison of arbitrary definitions of circadian time periods with those determined by wrist actigraphy in analysis of ABPM data. J Hum Hypertens 13:449–453

Ekelund U, Yngve A, Brage S, Westerterp K, Sjostrom M (2004) Body movement and physical activity energy expenditure in children and adolescents: how to adjust for differences in body size and age. Am J Clin Nutr 79:851–856

Epstein L, GcGowan C, Woodall K (1984) A behavioral observation system for free play activity in young overweight female children. Res Q Exerc Sport 55:180–183

Freedson P, Pober D (2005) Calibration of accelerometer output for children. Med Sci Sports Exerc 37:S523–S530

Fuchs RK, Bauer JJ, Snow CM (2001) Jumping improves hip and lumbar spine bone mass in prepubescent children: a randomized controlled trial. J Bone Miner Res 16:148–156

Janz KF (1994) Validation of the CSA accelerometer for assessing children's physical activity. Med Sci Sports Exerc 26:369–375

Kohl HW, Fulton JE, Caspersen CJ (2000) Assessment of physical activity among children and adolescents: a review and synthesis. Prev Med 31:S54–S76

McKenzie TL (1991) Observational measures of children's physical activity. J Sch Health 61:224–227

McKenzie TL, Sallis JF, Nader PR, Patterson TL, Elder JP, Berry CC, Rupp JW, Atkins CJ, Buono MJ, Nelson JA (1991) BEACHES: an observational system for assessing children's eating and physical activity behaviors and associated events. J Appl Behav Anal 24:141–151

Melanson EL Jr, Freedson PS (1995) Validity of the Computer Science and Applications, Inc. (CSA) activity monitor. Med Sci Sports Exerc 27:934–940

Metcalf BS, Curnow JS, Evans C, Voss LD, Wilkin TJ (2002) Technical reliability of the CSA activity monitor: the early bird study. Med Sci Sports Exerc 34:1533–1537

Pärkkä J, Ermes M, Korpipaa P, Mantyjarvi J, Peltola J, Korhonen I (2006) Activity classification using realistic data from wearable sensors. IEEE Trans Inf Technol Biomed 10:119–128

Pober DM, Staudenmayer J, Raphael C, Freedson PS (2006) Development of novel techniques to classify physical activity mode using accelerometers. Med Sci Sports Exerc 38:1626–1634

Puyau MR, Adolph AL, Vohra FA, Butte NF (2002) Validation and calibration of physical activity monitors in children. Obes Res 10(3):150–157

Ravi N, Dandekar N, Mysore P, Littman ML (2005) Activity recognition from accelerometer data. In: Proceedings of 20th national Conference on artificial intelligence (AAAI-05). American Association for Artificial Intelligence, pp 1541–1546

Riddoch CJ, Bo Andersen L, Wedderkopp N, Harro M, Klasson-Heggebo L, Sardinha LB, Cooper AR, Ekelund U (2004) Physical activity levels and patterns of 9- and 15-yr-old European children. Med Sci Sports Exerc 36:86–92

Tanaka C, Tanak S, Kawahara J, Midorikawa T (2007) Triaxial accelerometry for assessment of physical activity in young children. Obesity 15:1233–1241

Treuth MS, Schmitz K, Catellier DJ, McMurray RG, Murray DM, Almeida MJ, Going S, Norman JE, Pate R (2004) Defining accelerometer thresholds for activity intensities in adolescent girls. Med Sci Sports Exerc 36:1259–1266

Trost SG, Ward DS, Moorehead SM, Watson PD, Riner W, Burke JR (1998) Validity of the computer science and applications (CSA) activity monitor in children. Med Sci Sports Exerc 30:629–633

Trost SG, Pate RR, Freedson PS, Sallis JF, Taylor WC (2000) Using objective physical activity measures with youth: how many days of monitoring are needed? Med Sci Sports Exerc 32:426–431

Welk G, Schaben J, Morrow JA Jr (2004) Reliability of accelerometry-based activity monitors: a generaizibility study. Med Sci Sports Exerc 36:1637–1645